Recommendation-System Build Checklist — One Page

Fora Soft Learn

A video recommender answers two questions — what could they watch, and what first — and lives or dies on what it optimizes. Specify the two-stage funnel, the three relevance strategies, the cold-start fixes, and the one metric that keeps it honest before you build. Vendor capabilities move in 2026 — confirm live.

1 · THE TWO-STAGE FUNNEL (so it scales)

- ☐ **Candidate generation** — a cheap, wide net over the whole catalog; returns a few hundred (millions → hundreds).
- ☐ **Ranking** — the expensive model scores only that shortlist and orders the few dozen shown (hundreds → dozens).
- ☐ **Why split** — scoring the whole catalog per viewer per load cannot fit a ~200 ms home-screen budget.
- ☐ **Blend retrievers** — collaborative + content-based + editorial feed one shortlist; ranking sorts it out.

2 · THE THREE RELEVANCE STRATEGIES (plan for a hybrid)

- ☐ **Collaborative filtering** — "people like you watched this"; needs behavior; blind to brand-new titles.
- ☐ **Content-based** — "similar to what you watched"; runs on metadata; carries new titles but repeats.
- ☐ **Hybrid** — both, so each covers the other's blind spot; what nearly every real system ends up needing.
- ☐ **Metadata is the fuel** — content-based recs are only as good as a clean, structured catalog of metadata.

3 · THE COLD-START FIXES (the day it knows nothing)

- ☐ **New viewer** — onboarding taste picks, broad popularity, and context (country, device, time).
- ☐ **New title** — content metadata recommends it on day one; collaborative filtering cannot.
- ☐ **New platform** — lean on editorial + content-based first; shift to collaborative as watch data arrives.
- ☐ **Plan the warm-up** — the recommender you launch with is not the one you run a year later.

4 · MEASURE WATCH TIME, NOT CLICKS (the honest target)

- ☐ **Clicks lie** — ranking by click-through promotes "deceptive videos … ('clickbait')" (YouTube, RecSys 2016).
- ☐ **The honest metric** — expected watch time, completed sessions, and next-week return predict retention.
- ☐ **Prove it** — every change ships behind an experiment measured against retention, not taps.
- ☐ **Mind the boundaries** — viewing data has a privacy line; a perfect rec behind a spinner still loses the viewer.

THE ORDER OF OPERATIONS — TWO STAGES, A HYBRID, A WARM-UP PLAN, AND WATCH TIME AS THE TARGET

A recommendation system is mostly the plumbing around the model, not the model itself. Do it in order. First, split the work into two stages: a cheap candidate-generation step that narrows the whole catalog to a few hundred titles, then an expensive ranking step that scores only that shortlist — this is the only way to stay fast on a large catalog, and it is the pattern used by YouTube, Netflix, and Amazon. Second, plan for a hybrid: collaborative filtering finds the unexpected but is blind to new titles, content-based filtering carries new titles but repeats, so you will need both — and content-based recs are only as good as your metadata. Third, design for cold start from day one: warm new viewers with onboarding and popularity, carry new titles with content metadata, and launch a new platform on editorial plus content-based until behavior accumulates. Fourth, and most important, optimize ranking for expected watch time and return, never for clicks — a system tuned for clicks learns clickbait — and prove every change with an experiment tied to retention. Vendor capabilities and model architectures (the 2025 foundation-model shift, two-tower retrieval) move — re-verify before you plan against them.